DOCUMENT RESUME

ED 334 213 TM 016 635

AUTHOR	Hsiung, Tung-Hsing; Olejnik, Stephen
TITLE	Power of Pairwise Multiple Comparisons in the Unequal Variance Case.
PUB DATE	Apr 91
NOTE	34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Comparative Analysis; Equations (Mathematics); *Estimation (Mathematics); *Mathematical Models; Sample Size; Simulation
IDENTIFIERS	*Multiple Comparisons; *Paired Comparisons; Power (Statistics); Type I Errors

ABSTRACT

Using computer simulated data, the Type I error rate and statistical power were empirically estimated for several pairwise multiple comparison strategies for situations where population variances differ. Focus was on comparing modified Bonferroni procedures with Dunnett's solutions, and determining whether or not J. P. Shaffer's suggestion of using the omnibus test would work when population variances differed. Three factors were manipulated: sample size, variance heterogeneity, and pattern of population mean differences. Twenty-four different combinations of sample sizes and variance patterns were examined for the single factor four group design. The results indicate that all eight contrast procedures considered controlled the familywise Type I error rate under the nominal 0.05 level. In terms of statistical power, the Games-Howell procedure generally provided the greater power in identifying at least one significant difference. However, the magnitude of the any-pair power difference was very small. J. P. Shaffer's (1979) enhancements to the Bonferroni approach provided greater average power per contrast as well as the greatest power in identifying all significant pairwise differences. The results of the present study indicate that previous recommendations concerning the selection of a multiple comparison procedure when population variances differ should be reconsidered, and the adoption of the new strategies for multiple comparisons is recommended. Twelve data tables and a 22-item list of references are included. (Author/RLC)

POWER OF PAIRWISE MULTIPLE COMPARISONS IN THE UNEQUAL VARIANCE CASE

Tung-Hsing Hsiung

Stephen Olejnik

University of Georgia

Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. April, 1991.

ABSTRACT

Using computer simulated data, the Type I error rate and statistical power are empirically estimated for several pairwise multiple comparison strategies for situations where population variances differ. Twenty-four different combinations sample sizes and variance patterns are examined for the single factor four group design. The results indicate that all eight contrast procedures considered controlled the familywise Type I error rate under the nominal .05 level. In terms of statistical power, the Games-Howell procedure generally provided the greater power in identifying at least one significant difference. The magnitude of the any-pair power difference however was very small. Shaffer's (1979) enhancements to the Bonferroni approach provided greater average power per contrast as well as in identifying all significant pairwise differences. The results of the present study indicate that previous recommendations on the selection of a multiple comparison procedure when population variances differ should be reconsidered and the adoption of the new strategies for multiple comparisons is recommended.

Interest in contrast analysis has a long history in the statistical literature with work dating back to Fisher in 1935. Since then, the number of alternative procedures that have been developed has increased steadily, reflecting the continued interest in these techniques. Most of the procedures that have been developed and studied extensively, assume that the population variances to be equal. An excellent review of many of these alternatives can be found in Jaccard, Becker, and Wood (1984). A much more limited list of alternatives is available for situations were population variances differ (Kirk, 1990; Olejnik, 1990). Among the valid procedures for situations involving unequal population variances, the procedures developed by Games and Howell (1976) and Dunnett (1980) are the most frequently recommended. These procedures compute the test statistic by taking the ratio of the difference between group means to the standard error of the contrast, where the standard error is computed using separate variance estimates:

$$\overline{X_i} - \overline{X_j} \ / \ (s_i^2/n_i + s_j^2/n_j)^{1/2}.$$

Where: $\overline{X_i}$ and $\overline{X_j}$ are the sample means from groups i and j respectively; s_i and s_j are the standard deviations for group i and j respectively; and n_i and n_j are the sample sizes from groups i and j respectively. The procedures differ in

identifying the critical test statistic. Games-Howell (GH) uses $q_{\alpha,K,V_{ij}} / \sqrt{2}$. Where:

> q_α is the Studentized range distribution for the α centile; k is the number of groups in the family of contrasts; and V_{ij} are the approximate degrees of freedom from Satterthwaite (1947):

$$V_{ij} = \frac{(s_i^2/n_i + s_j^2/n_j)^2}{s_i^4/n_i^2(n_i-1) + s_j^4/n_j^2(n_j-1)} .$$

Dunnett (1980) suggested two solutions: one based Cochran's (1964) solution (C) to the Behrens-Fisher problem and uses $q_{\alpha,k,W_{ij}} / \sqrt{2}$. Where:

$$W_{ij} = \frac{q_{\alpha,k,n_i-1}\, s_i^2/n_i + q_{\alpha,k,n_j-1}\, s_j^2/n_j}{s_i^2/n_i + s_j^2/n_j}$$

and terms are defined as above.

The second solution (T3) suggested by Dunnett uses $A_{\alpha,C,V_{ij}}$ as the critical test statistic. Where:

> A_α is the Studentized maximum modulus distribution at the α centile; C is the number of contrasts in the family of comparisons; and V_{ij} is defined as above.

Data analysts differ in their recommendation as to the "best" approach. While several studies have shown the Games-Howell (GH) procedure to be robust to variance inequality (Keselman and Rogan, 1977; Games, Keselman, and Rogan, 1981)), Dunnett (1980) has provided some evidence to indicate that the Games-Howell procedure can be liberal in a limited set of situations. Wilcox (1987) in studying

factorial designs also found the Games-Howell procedure to be liberal when sample sizes are small.

In discussing statistical power, Dunnett considered the width of confidence intervals to conclude that the T3 method would provide a narrower interval width than the C approach when limited degrees of freedom were available but the C method would provide a narrower interval range if sample size was large. Other indicators of statistical power such as: a) the probability of identifying at least one significant contrast, any-pair power, b) the average power per-contrast, per-pair power, or c) the probability of identifying all significant contrasts, all-pairs power (Ramsey, 1978, 1981; Einot & Gabriel, 1975) were not considered.

A popular multiple comparison procedure not considered by Dunnett is the Bonferroni adjusted t-test (B). When population variances differ the computed test statistic is calculated as above but uses $t_{\alpha,V_{ij}}$ as the critical test statistic. Where $t_{\alpha'}$ is the Student t distribution at the α' centile; and V_{ij} is defined above. This approach is generally considered relevant when a subset of all possible contrasts are of interest and can specified before data are collected. But the procedure is not limited to those situations and can be used to test all contrasts providing that an appropriate adjustment is made to control Type I error rate familywise. The Bonferroni approach for

controlling the overall Type I error rate is to divide the desired familywise significance level equally among the c contrasts. Each contrast is then tested for statistical significance using an adjusted alpha level equal to $\alpha' = \alpha/c$.

Recently, several researchers (Holm, 1979; Holland and Copenhaver, 1988) have developed strategies to improve the statistical power for the Bonferroni technique by modifying the criteria used for statistical significance. Holm (H) suggested that the adjusted significance level be based on rank ordering of the mean differences and the significance level for a contrast be based on its ranking (r_i), where the largest mean difference is given a rank of 1 and the smallest mean difference is given a rank of c. The significance level for the r_i contrast is set equal to: $\alpha_i' = \alpha/(c-r_i+1)$. Alternatively, Holland and Copenhaver (HC) (1987) suggested setting the significance for a contrast equal to $\alpha_i' = 1-(1-\alpha)^{[1/(c-r_i+1)]}$.

When all pairwise contrasts are of interest, the Bonferroni and Holm criteria for significance for the largest contrast is the same but for the remaining c-1 contrasts Holm's criteria (α_i', i= 2 to c) is consistently larger, thus facilitating the rejection of the null hypothesis. Holland and Copenhaver's criteria (α_i') for the largest contrast is always larger than the Bonferroni and Holm's procedures and slightly larger than Holm's criteria for the next c-2 contrasts but is identical for the smallest

contrast where both criteria are equal to α. Thus there is a slight power advantage associated with Holland and Copenhaver's approach. Shaffer (1979, 1986) suggested modifications to multiple-range procedures like Holm's adjustment to the Bonferroni approach that would lead to a further enhancement in statistical power. Shaffer pointed out that the rejection of one contrast in a family of contrasts often has a logical consequence for some or all of the remaining contrasts of interest. For example, if all pairwise contrasts among three populations are of interest and one contrast is rejected, then at least one other contrast must also be rejected. By taking into consideration the logical interrelation among the contrasts the statistical power can be improved. Shaffer showed that the enhancement in statistical power does not come as a consequence of an increased risk of the overall familywise significance level. When examining all pairwise contrasts Shaffer (S) suggested the largest contrast can be tested for significance with the criteria set equal to α/c and the remaing c-1 contrasts set equal to α/t, where t is set equal to the number of the remaining contrasts that could possibly be true. Seaman, Levin, Serlin and Franke (1990) have developed an algorithm for determining t when all pairwise contrasts are of interest and provide a table of t values when all pairwise contrasts are of interest among 3, 4, or 5 populations. Following this procedure the criteria for

significance for the largest contrast equals $\alpha' = \alpha/c$, the same criteria used in the Bonferroni and Holm's procedures. The smallest contrast is tested using the same criteria as Holm and Holland's. Shaffer's criteria is always as large or larger than Holm's criteria but can be smaller than Holland and Copenhaver's criteria. Thus Shaffer's procedure can never be less powerful than Holm's procedure but could be less powerful than Holland and Copenhaver's.

An alternative to focusing exclusively on the contrasts, Shaffer (S1) (1979) showed that a preliminary omnibus test could be used in lieu of the contrast between the largest and smallest means and given that the overall test was rejected pairwise contrasts could proceed as with S except the largest contrast would use as the significance level the same criteria as the second largest contrast. Thus if the omnibus test was significant S1 would be more powerful than S only for the largest contrast.

In an extensive examination of these modifications to the Bonferroni procedure in situations where population variances were equal Seaman, Levin, Serlin and Franke (1990) recommended Shaffer's (S1) modification of Holm's procedure using the preliminary ANOVA F-test. Although it was not the most powerful procedure from those considered, it was the most powerful procedure studied that was reasonably easy to implement. The most powerful procedure examined in that study was a strategy which examined all possible partitions

of k means. This procedure while controlling the familywise error rate at the nominal level and maximizing the statistical power is computationally intensive and thus not likely to be accepted by the applied researcher without supporting computer software.

These enhancements to the Bonferroni adjustment have not been applied to situations involving unequal population variances and have not been compared to the Games-Howell, the T3 or the C techniques. When population variance differ ordering the contrasts by mean differences is not meaningful but the ranking necessary for the enhancement of the Bonferroni approach can be achieved by ordering the observed probability values from the smallest p-value to the largest. Contrasts can be rejected when $p_i<\alpha_i'$.

PURPOSE

The purpose of the present study is compare the familywise Type I error rate and statistical power of the Games-Howell (GH), Dunnett T3, Dunnett C, Bonferroni (B), Holm (H), Holland-Copenhaver (HC), Shaffer, S, and Shaffer S1 for pairwise contrasts when population variances differed in balanced and unbalanced one factor designs. Since there is some evidence to indicate that the Games-Howell procedure can be liberal we were particularly interested in comparing the modified Bonferroni procedures with the Dunnett's solutions. We were also interested in determining whether

Shaffer's suggestion of using the omnibus test (S1) would work when population variances differed. When population variances differ the omnibus suggested by Welch (1951) and Brown and Forsythe (1974) are often recommended. These procedures however can be liberal (Wilcox, Charlin, and Thompson, 1986) when population variances differ greatly. Thus Shaffer's preliminary F-test solution may not be appropriate when variances differ. In the present study we use Brown and Forsythe's (1974) adjustment to the parametric F-test when population variances differ. The test statistic is computed as:

$$\frac{\Sigma n_a(X_{.a}-X_{..})^2}{\Sigma(1-n_a/N)s_a^2}$$

The critical test statistic is found in the central F-distribution with a-1 and f degrees of freedom. Where f is computed as:

$$f = \{\Sigma[c_a^2/(n_a-1)]\}^{-1}$$

$$C_a = \frac{(1-n_a/N)s_a^2}{\Sigma(1-n_a/N)s_a^2}$$

METHOD

The study is carried out using computer generated data. Three factors were manipulated: 1) sample size, 2) variance heterogeneity, and pattern of population mean differences. The study is limited to a single factor four group design and it is assumed the researcher is interested in all

pairwise contrasts (c=6). Table 1 summarizes the patterns of sample size and variance conditions studied. For each of

Insert Table 1 about here

these patterns we considered situations where all population means were equal to study Type I error rates. To study statistical power we considered two patterns of mean differences: (1) $\mu_1 > \mu_2 = \mu_3 = \mu_4$ and (2) $\mu_1 > \mu_2 > \mu_3 > \mu_4$. For the second pattern we included two levels of mean differences.

Data are generated using SAS-Proc Matrix (1985). Under each of the conditions outlined above data are generated for the following linear model: $Y_{ia} = Mu + ES_{.a} + E_{ia}$. Where Mu is the grand mean set equal to 10 for the study; ES is the effect size equalling 0 for the null condition. For pattern 1 μ_1 was set equal to 1 and the other groups were set equal to 0. For pattern 2 the four means were set equal to 1.5, 1.0, .5, 0 and 1.0, .5, .25, 0 for groups one to four respectively. The random error component E_{ia} was normally distributed with a mean equalling 0 and variance set equal to the patterns presented in Table 1. To generate the random error component the SAS normal random generating function, RANOR, is used. For each condition studied 5000 replications were generated and the proportion of times the contrasts were rejected at the .05 level is recorded. To evaluate the results under the complete null and partial

null conditions, if the proportion of null hypotheses rejected experimentwise exceeds .056 (two standard errors above the .05 level) it is concluded that the procedure is liberal. To evaluate power, difference in rejection rates for any of the three definitions of power greater than .03 will be interpreted as of practical importance.

RESULTS

Type I Error Rates.

The Type I error rates for the eight pairwise multiple comparison procedures and the omnibus parametric and Brown and Forsythe's (BF) adjusted F-test for the 24 patterns are reported in Table 2. Both the parametric and adjusted

Insert Table 2 about here

omnibus F-tests have Type I error rates that are seriously affected by the unequal variance and unequal sample size patterns. Even with relatively small variance differences BF's adjustment did not control the overall Type I error rate (see patterns 4, 7, 19, 20). The problem is more severe when variance differ greatly (see patterns 12, 13, 14, 15, 16, 23, 24). These results are consistent with the findings reported by Wilcox, Charlin and Thompson (1986).

All of the multiple comparison procedures considered here however, controlled the familywise Type I error rate under the nominal level. Thus although the omnibus Brown-

Forsythe adjusted F-test had Type I error rates greater than the nominal level, the Type I error rate of Shaffer's modification of the Bonferroni following this omnibus test never exceeded the nominal level. The empirical Type I error rate for S1 appears to be in the same magnitude as Dunnett's T3 procedure.

The empirical Type I error rate for the Games-Howell procedure ranged between .040 and .055. Thus for the four group design the Games-Howell procedure does not appear to be liberal. This result is consistent with the findings reported by Dunnett (1980) for the four group design he considered.

As was found in Dunnett's (1980) study Cochran's solution to the Behrens-Fisher problem resulted in the smallest rejection rate when sample sizes were small. With small samples then the C procedure appears to be very conservative.

Finally, the Bonferroni, Holm's modification, and Shaffer's (S) procedures all had the same Type I error rate. This result was expected since all three procedures use the same criteria for rejecting the contrast with the largest p-value.

For the pattern of means where $\mu 1 > \mu 2 = \mu 3 = \mu 4$ we also examined the Type I error rate for the contrasts involving equal population means ($\mu 2 = \mu 3 = \mu 4$). All eight of the procedures studied had partial Type I error rates

less than .05. That is for the three null contrasts ($\mu2$ = $\mu3$, $\mu2$ = $\mu3$ and $\mu3$ = $\mu4$), none of the procedures rejected at least one of these contrasts more than 5% of the time.

Statistical Power

Based on the two non-null patterns, $\mu_1>\mu_2=\mu_3=\mu_4$ and $\mu_1>\mu_2>\mu_3>\mu_4$ the any pair power, per-pair power and all pair power were estimated for the 24 patterns identified in Table 1. Some of the results for the any-pair and all-pair power were discarded because the magnitude of the mean differences studied gave the any-pair power estimates close to 1 for all of the procedures studied and close to 0 for all procedures when all- pair power was estimated. The results are presented below by definition of power since the conclusions vary as a function of the definition.

Any-Pair Power.

Table 3 presents a rank ordering of the any-pair power estimates for the eight multiple comparison procedures when population means were 1, 0, 0, 0 and 1, .5, .25, 0. For the any-pair power definition the Bonferroni, Holm and Shaffer (S) have the same power estimate so the table only includes a column for S. A comparison of the five highest ranked procedures is summarized in Table 4. Values in the table indicate the proportion of conditions that the procedure providing the column label had any-pair power greater than a procedure identified by the row label. For example the GH (Games-Howell) procedure always had greater any-pair power

than the T3 procedure. These comparisons were made across all three population mean difference conditions. These

Insert Tables 3 and 4 about here

results indicate that for the 24 patterns studied, the Games-Howell procedure generally provided the most statistical power for identifying at least one pairwise difference. Shaffer's procedure preceded by the omnibus adjusted F-test however did provide the greater power when the group with the largest mean also had the largest variance. Tables 5 and 6 provide estimates of the magnitude of the any-pair power differences between selected contrast procedures. These results

Insert Tables 5 and 6 about here

indicate that the power differences were generally small with over half being less than three percent.

Per-Pair Power. Table 7 summarizes the ranking ordering of the eight contrast procedures using the average power per contrast definition. Table 8 reports the proportion of conditions that a procedure identified by the

Insert Table 7 and 8 about here

column heading had greater power than a procedure identified by the row label. These results indicate that Shaffer's

procedures generally alternate in identifying the greatest average power per contrast. Tables 9 and 10 indicate the magnitude of the power difference. These results indicate

---

Insert Table 9 and 10 about here

---

that in general the modifications to the Bonferroni procedure provide a substantial increase in average power per contrast than either the Games-Howell or Dunnett's T3 procedures. The magnitude of the difference in power between Shaffer's procedures was small but generally favored S1.

All-Pair Power. To identify all significant differences among the four populations the results indicate that Shaffer's modifications to the Bonferroni strategy provided the greatest power. Table 11 presents the rank ordering of the contrast procedures based on the all-pair

---

Insert Table 11 about here

---

definition of power. Shaffer's S1 procedure using the omnibus adjusted F-test generally provided the most powerful approach but Shaffer's alternative S was consistently ranked second. Table 12 provides estimates of the magnitude of the power differences and the results

---

Insert Table 12 about here

---

indicate a very small difference in all-pairs power between the two procedures suggested by Shaffer. Holland's procedure was generally ranked third and also consistently offered greater power than Games-Howell or Dunnett's T3 procedures.

CONCLUSIONS

The present study only considered the single factor four group design with a limited number of sample size and variance combinations. As a result broad generalizations cannot be made. However the conditions that were studied included many of the situations frequently encountered by the applied researcher. The results of this study are probably best viewed as an indication of the relative merits of the alternative approaches to multiple comparisons when variances differ. With these limitations in mind, the following conclusions seem justified:

1) When all pairwise contrasts among four populations are of interest and the nominal familywise Type I error rate is set at .05, all eight of the multiple comparison procedures considered in this study had empirical Type I error rates that did not exceed two standard errors of the nominal significance level.

2) Although the Brown-Forsythe omnibus adjusted F-test was shown to be liberal for many combinations of sample size and variance heterogeneity, the Shaffer's enhancement to the Holm's modification of the Bonferroni procedure, which relies on the results of the omnibus test, is not liberal at least for the conditions we studied.

3) The identification of the most powerful multiple comparison procedure for the unequal variance case depends on the definition of power. To identify at least one significant difference the Games-Howell procedure typically will provide the most sensitive test. However the difference in power between Shaffer's procedure using the omnibus adjusted F-test and the Games-Howell procedure is very small.

4) To maximize the average power per contrast or to identify all significant pairwise differences, either of Shaffer's procedures can be recommended.

5) Dunnett's alternatives generally had lower power across all definitions of power than the Games-Howell procedure or any of the modifications of the Bonferroni approach.

REFERENCES

Brown, M. & Forsythe, A. (1974). The small sample behavior of some statistics which test the equality of several means. Technometrics, 9, 129-132.

Cochran, W. G. (1964). Approximate significance levels of the Behrens-Fisher test. Biometrics, 20, 191-195.

Dunnett, C. W. (1980). Pairwise multiple comparisons in the unequal variance case. Journal of the American Statistical Association, 75, 796-800.

Einot, I., & Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. Journal of the American Statistical Association, 70, 574-583.

Games, P. A. & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal N's and /or variances: A monte carlo study. Journal of Educational Statistics, 1, 113-125.

Games, P. A., Keselman, H. J.,& Rogan, J. C. (1981). Simultaneous pairwise multiple comparison procedures for means when sample sizes are unequal. Psychological Bulletin, 90, 594-598.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.

Holland, B. S. & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. Psychological Bulletin, 104, 145-149.

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise multiple comparison procedures: A review. Psychological Bulletin, 96, 589-596.

Keselman, H. J., & Rogan, J. C. (1978). A comparison of the notified-Tukey and Scheffe methods of multiple comparisons for pairwise contrasts. Journal of the American Statistical Association, 73, 47-52.

Kirk, R. (April, 1990). Multiple comparisons: From Fisher to the future. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Olejnik, S. & Lee, J. (April, 1990). Multiple comparison procedures when population variances differ. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA

Ramsey, P. H. (1978). Power differences between pairwise multiple comparisons. Journal of the American Statistical Association, 73, 479-485.

Ramsey, P. H. (1981). Power of univariate pairwise multiple comparison procedures. Psychological Bulletin, 90, 352-366.

SAS Users Guide: Basics. Version 5 Edition (1985). Cary, NC: SAS Institute, Inc.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. Biometrics, 2, 110-114.

Seaman, M. A., Levin, J., R., Serlin, R. C., & Franke, T. M. (April, 1990). A modified algorithm and an examination of power for Peritz's pairwise multiple-comparison procedures. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Shaffer, J. P. (1979). Comparison of means: An F followed by a modified multiple range procedure. Journal of Educational Statistics, 4, 14-23.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. Journal of the American Statistical Association, 81, 826-831.

Welch, B. L. (1947). The generalization of student's problems when several different population variances are involved. Biometrika, 34, 28-35.

Wilcox, R. R., Charlin, V. & Thompson, K. (1986). New monte carlo results on the robustness of the ANOVA F, W and F^* statistics. Communications in Statistics-Simulation and Computation, 15, 933-944.

Wilcox, R. R. (1987). Within row pairwise comparisons in a two-way ANOVA. Communications in Statistics-Simulation, 16, 939-955.

Table 1.

Patterns of sample size and variance differences examined.

n_1	n_2	n_3	n_4	σ_1^2	σ_2^2	σ_3^2	σ_4^2	Pattern
7	7	7	7	1	1	1	1	1
22	22	22	22	1	1	1	1	2
14	18	24	28	1	1	1	1	3
7	7	7	7	.5	1.5	1.5	1.5	4
22	22	22	22	.5	1.5	1.5	1.5	5
14	18	24	28	.5	1.5	1.5	1.5	6
28	36	48	56	.5	1.5	1.5	1.5	7
7	7	7	7	.6	.8	1.0	1.2	8
22	22	22	22	.6	.8	1.0	1.2	9
14	18	24	28	.6	.8	1.0	1.2	10
28	36	48	56	.6	.8	1.0	1.2	11
7	7	7	7	.1	.4	.8	1.6	12
22	22	22	22	.1	.4	.8	1.6	13
44	44	44	44	.1	.4	.8	1.6	14
14	18	24	28	.1	.4	.8	1.6	15
28	36	48	56	.1	.4	.8	1.6	16
7	7	7	7	2.2	.6	.6	.6	17
22	22	22	22	2.2	.6	.6	.6	18
14	18	24	28	2.2	.6	.6	.6	19
28	36	48	56	2.2	.6	.6	.6	20
14	18	24	28	1.2	1.0	.8	.6	21
28	36	48	56	1.2	1.0	.8	.6	22
14	18	24	28	1.6	.8	.4	.1	23
28	36	48	56	1.6	.8	.4	.1	24

Table 2.

Type I errors for ANOVA F-ratio and alternative multiple comparison procedures for various sample size and variance patterns.

Pattern	GH	T3	C	B	H	HL	S	S1	F	F1
1	053	042	023	033	033	032	033	034	050	047
2	052	043	041	042	042	043	042	046	050	050
3	053	045	040	042	042	043	042	046	051	052
4	053	042	026	033	033	035	033	037	070	057
5	052	043	040	040	040	041	040	046	054	053
6	051	041	038	039	039	040	039	044	043	056
7	054	044	049	045	045	046	045	051	045	058
8	055	043	024	032	032	033	032	039	052	046
9	046	039	037	037	037	038	037	041	047	046
10	047	040	035	037	037	039	037	042	037	050
11	051	040	043	041	041	042	041	048	040	051
12	050	042	032	034	034	035	034	036	073	059
13	040	034	035	032	032	032	032	035	066	061
14	045	036	042	036	036	037	036	040	070	066
15	053	042	042	039	039	040	039	046	039	066
16	044	036	040	036	036	037	036	043	034	067
17	049	041	025	032	032	034	032	038	066	055
18	049	041	039	039	039	040	039	044	052	051
19	053	042	040	039	039	040	039	039	110	067
20	048	039	042	038	038	038	038	039	106	068
21	047	040	038	038	038	039	038	043	067	051
22	052	042	044	041	041	042	041	050	072	056
23	047	040	041	037	037	037	037	038	132	064
24	045	037	040	036	036	036	036	041	127	068

Table 3.

Rank orderings of contrast procedures for any pair power.

	$\mu_1>\mu_2=\mu_3=\mu_4$ (1.0,0,0,0)						$\mu_1>\mu_2>\mu_3>\mu_4$ (1.0,.5,.25,0)					
Pattern	Gh	T3	C	HL	S	S1	GH	T3	C	HL	S	S1
1	1	3	6	4	5	2	1	3	6	4	5	2
2	2	3	5.5	4	5.5	1	2	3	6	4	5	1
3	2	3	6	4	5	1	2	3	6	4	5	1
4	1	2	6	3	4	5	1	2	6	3	4	5
5	1	3	2	5	4	6	1	2	3	4	5	6
6	1	2	5	3	4	6	1	2	5	3	4	6
7	1	5	3	3	3	6	1	5	2	3	4	6
8	1	2	6	4	5	3	1	3	6	4	5	2
9	1	2	6	4	3	5	1	3	5	4	6	2
10	1	2	5	4	3	6	1	3	6	4	5	2
11			NA				1	6	2	4.5	4.5	3
12	1	2	3	4	5	6	1	2	3	4	5	6
13			NA				1	2	3	4	5	6
14			NA						NA			
15			NA						NA			
16			NA						NA			
17	2	3	6	4	5	1	1	3	6	4	5	2
18	2	4	3	5	6	1	2	3	4	5	6	1
19	2	4	3	5	6	1	1	4	3	5	6	2
20	2	4	3	6	5	1	2	4	3	5	6	1
21	2	3	4	6	5	1	2	4	3	5	6	1
22	2	4	3	5	6	1	2	4	3	5	6	1
23	2	4	3	5	6	1	1	4	3	5	6	2
24	2	4	3	5	6	1	1	5	3	4	6	2

Table 4

Percent of time that procedures identified by the column label had any-pair power greater than an alternative identified by the row label.

	GH	T3	Holland	Shaffer	Shaffer1
GH		0	.02	0	.45
T3	1.0		.07	.04	.58
Holland	.98	.93		.05	.69
Shaffer	1.0	.96	.89		.69
Shaffer1	.55	.40	.31	.31	

* Total of 54 conditions

Table 5

Any-pair Power difference between two procedures. $\mu 1 > \mu 2 = \mu 3 = \mu 4$ (1>0=0=0).

A = GH - T3; B = GH - Holland; C = GH - Shaffer; D = GH - Shaffer1; E = T3 - Holland; F = T3 - Shaffer; G = T3 - Shaffer1; H = Holland - Shaffer; I = Holland - Shaffer1; J = Shaffer - Shaffer1.

Pattern	A	B	C	D	E	F	G	H	I	J
1	.04	.07	.07	+	.03	.03	+	+	-.06	-.06
2	+	+	+	-	+	+	-	+	-	-
3	+	.03	.03	-	+	+	-.03	+	-.04	-.04
4	.04	.08	.09	.11	.04	.04	.07	+	.03	+
5	+	+	+	.04	+	+	+	+	+	+
6	+	+	+	.13	+	+	.11	+	.11	.11
7	+	+	+	+	-	-	+	0	+	+
8	.04	.08	.09	.07	.04	.04	+	+	-	-
9	+	+	+	+	+	+	+	+	+	+
10	+	+	+	.03	+	+	+	+	+	+
11	0	0	0	0	0	0	0	0	0	0
12	.04	.08	.09	.28	.05	.05	.24	+	.20	.20
13					NA					
14					NA					
15					NA					
16					NA					
17	+	.05	.05	-	+	+	.06	+	-.07	-.08
18	+	.04	.04	-.06	+	+	-.09	+	-.09	-.10
19	.03	.05	.05	.07	+	+	-.10	+	-.11	-.11
20	+	+	+	-.05	+	+	-.06	+	-.07	-.07
21	.03	.04	.04	-.04	+	+	-.08	+	-.07	-.07
22	+	+	+	-	+	+	-	+	-	-
23	.05	.06	.07	-.06	+	+	-.10	-	-.12	-.12
24	+	+	+	+	+	+	-.03	+	-.03	-.04

+ = positive difference but less than .03;
- = negative difference but large than -.03;
NA = not availabe, all power value are equal.

Table 6

Any-pair Power difference between two procedures. $\mu 1 > \mu 2 > \mu 3 > \mu 4$ $(1 > .5 > .25 > 0)$.

A = GH - T3; B = GH - Holland; C = GH - Shaffer; D = GH - Shaffer1; E = T3 - Holland; F = T3 - Shaffer; G = T3 - Shaffer1; H = Holland - Shaffer; I = Holland - Shaffer1; J = Shaffer - Shaffer1.

Pattern	A	B	C	D	E	F	G	H	I	J
1	.03	.06	.07	+	+	.03	-	+	-.04	-.04
2	+	.03	.038	-	+	+	-.03	+	-.04	-.04
3	.03	.04	.048	-	+	+	-.04	+	-.05	-.06
4	.03	.06	.07	.08	+	.03	.05	+	+	+
5	+	+	.03	.08	+	+	.05	+	.05	.05
6	+	.03	.03	.15	+	+	.12	+	.12	.11
7	+	+	+	.03	+	+	.03	+	.03	.03
8	.04	.07	.07	.04	.03	.04	-	+	-.03	-.04
9	+	+	+	+	+	+	-	+	-	-
10	+	+	.03	+	+	+	-	+	-	-
11	+	+	+	+	-	-	-	0	-	-
12	.04	.09	.10	.13	.05	.06	.09	+	.04	+
13	+	+	+	.04	+	+	.03	+	.03	.03
14					NA					
15	+	+	+	.07	+	+	.07	+	.07	.07
16					NA					
17	+	.04	.05	+	+	+	-	+	-	-
18	.03	.03	.04	-	+	+	-.06	+	-.06	-.07
19	.04	.05	.05	+	+	+	-	+	-.03	-.03
20	+	+	+	-	+	+	-	+	-	-
21	.04	.05	.05	-	+	+	-.06	+	-.07	-.08
22	+	+	+	-	+	+	-	+	-	-
23	.04	.06	.06	+	+	+	-.03	+	-.05	-.05
24	+	+	+	+	-	+	-	+	-	-

+ = positive difference but less than .03;
- = negative difference but large than -.03;
NA = not available, all power value are equal.

Table 7.

<u>Rank orderings of contrast procedures for per-pair power</u>.

	$\mu_1>\mu_2=\mu_3=\mu_4$ (1.0,0,0,0)								$\mu_1>\mu_2>\mu_3>\mu_4$ (1.0,.5,.25,0)							
Pattern	GH	T3	C	B	H	HL	S	S1	GH	T3	C	B	H	HL	S	S1
1	2	4	8	7	6	5	3	1	2	4	8	7	6	5	3	1
2	5	6	8	7	4	3	2	1	3	6	8	7	5	4	2	1
3	3	6	8	7	5	4	2	1	3	6	7	8	5	4	2	1
4	1	3	7	8	6	5	2	4	1	3	8	7	6	5	2	4
5	5	6	7	8	4	3	1	2	3	6	7	8	5	4	1	2
6	2	6	8	7	4	3	1	5	2	6	8	7	4	3	1	5
7	5	7	6	8	4	3	1	2	5	8	6	7	4	3	1	2
8	1	4	8	7	6	5	3	2	2	4	8	7	6	5	3	1
9	5	6	8	7	4	3	2	1	5	6	8	7	4	3	2	1
10	4.5	6	8	7	4.5	3	1	2	5	6	8	7	4	3	2	1
11	5	7	6	8	4	3	1	2	5	8	6	7	4	3	2	1
12	1	3	5	7	6	4	2	8	1	3	7	8	6	5	2	4
13	5	7	6	8	4	3	1	2	5	7	6	8	4	3	1	2
14	5	7	6	8	3.5	3.5	1.5	1.5	5	8	6	7	3	1	3	3
15	5	6	7	8	4	3	1	2	5	6	7	8	3	2	1	4
16	5	7.5	6	7.5	4	3	1.5	1.5	5	8	6	7	3	2	1	4
17	2	4	8	7	6	5	3	1	2	4	8	7	6	5	3	1
18	3	8	6	7	5	4	2	1	4	6	7	8	5	3	2	1
19	3	7	6	8	5	4	2	1	3	7	6	8	5	4	2	1
20	3	7	6	8	5	4	2	1	5	7	6	8	4	3	2	1
21	3	6	7	8	5	4	2	1	3	6	7	8	5	4	2	1
22	5	7	6	8	4	3	2	1	5	7	6	8	4	3	2	1
23	3	7	5	8	6	4	2	1	3	7	6	8	5	4	2	1
24	4	7	6	8	5	3	2	1	5	7	6	8	4	3	2	1

Table 8

Percent of time that procedures identified by the column label had per-pair power greater than an alternative identified by the row label.

	GH	T3	Holland	Shaffer	Shaffer1
GH		0	.64	.82	.88
T3	1.0		.78	1.0	.97
Holland	.33	.22		.90	.83
Shaffer	.18	0	.08		.56
Shaffer1	.10	.03	.15	.33	

* Total of 72 conditions

Table 9

Per-pair Power difference between two procedures. $\mu1 > \mu2 = \mu3 = \mu4$ (1>0=0=0).

A = GH - T3; B = GH - Holland; C = GH - Shaffer; D = GH - Shaffer1; E = T3 - Holland; F = T3 - Shaffer; G = T3 - Shaffer1; H = Holland - Shaffer; I = Holland - Shaffer1; J = Shaffer - Shaffer1.

Pattern	A	B	C	D	E	F	G	H	I	J
1	+	.03	+	-	+	-	-.03	-	-.04	-
2	+	-	-.04	-.05	-	-.06	-.08	-.03	-.05	-
3	.03	+	-	-	-	-.06	-.08	-.04	-.06	-
4	+	.04	+	.08	+	-	.05	-	.04	.05
5	+	-	-.04	-.04	-	-.04	-.03	-.04	-.03	+
6	+	-	-.04	+	-.03	-.06	-	-.03	+	.05
7	+	-	-	-	-	-	-	-	-	+
8	+	.04	+	+	+	-	-	-	-.03	-
9	+	-	-	-	-	-.05	-.05	-	-	-
10	+	-	-.03	-	-	-.06	-.05	-	-	+
11	+	-	-	-	-	-	-	-	-	+
12	.04	.06	+	.09	+	-	.06	-.04	.03	.08
13	+	-	-	-	-	-	-	-	-	+
14	+	-	-	-	-	-	-	-	-	0
15	+	-	-.03	-.03	-.03	-.05	-.05	-	-	+
16	+	-	-	-	-	-	-	-	-	0
17	+	+	-	-	+	-	-.04	-	-.04	-.03
18	.03	+	-	-.07	-	-.05	-.10	-.03	-.07	-.40
19	.03	-	-	-.06	-	-.04	-.09	-	-.08	-.05
20	+	+	-	-.05	-	-	-.08	-	-.06	-.03
21	.03	+	-	-.06	-	-.05	-.09	-.03	-.07	-.04
22	+	-	-	-	-	-.03	-.04	-	-	-
23	.04	+	-	-.06	-	-.04	-.10	-	-.09	-.06
24	+	-	-	-.04	-	-.04	-.05	-	-.04	-

+ = positive difference but less than .03;
- = negative difference but large than -.03;
NA = not available, all power value are equal.

Table 10

Per-pair Power difference between two procedures. $\mu1 > \mu2 = \mu3 = \mu4$ (1>.5>.25>0).

A = GH - T3; B = GH - Holland; C = GH - Shaffer; D = GH - Shaffer1; E = T3 - Holland; F = T3 - Shaffer; G = T3 - Shaffer1; H = Holland - Shaffer; I = Holland - Shaffer1; J = Shaffer - Shaffer1.

Pattern	A	B	C	D	E	F	G	H	I	J
1	+	+	+	-	+	-	-	-	-	-
2	+	-	-	-.03	-	-.04	-.05	-	-.03	-
3	+	+	-	-.03	-	-.03	-.05	-	-.03	-
4	+	+	+	+	+	-	+	-	-	+
5	+	+	-	-	-	-.04	-	-	-	+
6	+	+	-	+	-	-	-	-	+	+
7	+	-	-.04	-.03	-.04	-.06	-.05	-	-	+
8	+	+	+	-	+	-	-	-	-	+
9	+	-	-	-	-.04	-	-.05	-.05	-	-
10	+	-	-	-	-	-.04	-.04	-	-	-
11	+	-.03	-.04	-.04	-.05	-.06	-.06	-	-	-
12	+	+	+	+	+	-	+	-	-	+
13	+	-	-.04	-.03	-.04	-.06	-.05	-	-	+
14	+	-.04	-.03	-.03	-.05	-.05	-.05	+	+	0
15	+	-	-.04	-	-.04	-.06	-.03	-	+	+
16	+	-.04	-.04	-.04	-.06	-.06	-.05	+	+	+
17	+	+	+	-	+	-	-	-	-	-
18	+	-	-.04	-	+	-.04	-.05	-	-.03	-
19	+	+	-	-	-	-.03	-.04	-	-	-
20	+	-	-.04	-.05	-.05	-.07	-.07	-	-	-
21	+	+	-	-.03	-	-.04	-.05	-	-.04	-
22	+	-	-.04	-.05	-.05	-.07	-.07	-	-	-
23	+	+	-	-.03	-	-.04	-.05	-	-.03	+
24	+	-.04	-.06	-.06	-.07	-.08	-.08	-	-	-

+ = positive difference but less than .03;
- = negative difference but large than -.03;
NA = not available, all power value are equal.

Table 11.

Rank ordierings of the eight contrast procedures based on all-pair power when $\mu_1>\mu_2=\mu_3=\mu_4$.

Pattern	GH	T3	C	B	H	HL	S	S1
1	3	6	8	7	4	5	2	1
2	3	6	8	7	5	4	2	1
3	3	6	8	7	5	4	2	1
4	3	6	7	8	4.5	4.5	2	1
5	5	6	7	8	4	3	2	1
6	5	7	8	6	4	3	1	2
7	5	7	6	8	4	3	1	2
8	3	6	8	7	5	4	2	1
9	5	6	7	8	4	3	2	1
10	5	6	8	7	4	3	1	2
11	5	7	6	8	4	3	1.5	1.5
12	3	6	7	8	5	4	2	1
13	5	7	6	8	4	3	1.5	1.5
14	5	7	6	8	3.5	3.5	1.5	1.5
15	5	7	6	8	4	3	1	2
16	5	7.5	6	7.5	3.5	3.5	1.5	1.5
17	3	6	8	7	5	4	2	1
18	5	6	7	8	4	3	2	1
19	5	7	6	8	4	3	2	1
20	3	7	6	8	4	3	2	1
21	5	6	8	7	4	3	2	1
22	5	7	6	8	4	3	2	1
23	5	6	7	8	4	3	2	1
24	5	7	6	8	4	3	2	1

Table 12

All-pair Power difference between two procedures. $\mu 1 > \mu 2 = \mu 3 = \mu 4$ (1>0=0=0).

A = GH - T3; B = GH - Holland; C = GH - Shaffer; D = GH - Shaffer1; E = T3 - Holland; F = T3 - Shaffer; G = T3 - Shaffer1; H = Holland - Shaffer; I = Holland - Shaffer1; J = Shaffer - Shaffer1.

Pattern	A	B	C	D	E	F	G	H	I	J
1	+	+	-	-	-	-	-	-	-	-
2	.03	.03	-.08	-.08	-.07	-.11	-.11	-.05	-.05	-
3	.03	-	-.08	-.08	-.06	-.10	-.11	-.05	-.05	-
4	+	-	-	-	-	-	-	-	-	-
5	.03	-.03	-.09	-.09	-.07	-.13	-.13	-.06	-.06	-
6	.04	-.03	-.07	-.07	-.07	-.11	-.10	-.04	-.04	+
7	+	-	-.03	-.03	-.03	-.05	-.05	-	-	+
8	+	+	-	-	-	-	-	-	-	-
9	.04	-	-.06	-.06	-.06	-.10	-.10	-.04	-.04	-
10	.03	-.03	-.08	-.08	-.06	-.11	-.11	-.05	-.04	+
11	+	-	-	-	-	-	-	-	-	0
12	+	+	-	-	-	-	-	-	-	-
13	.03	-	-.05	-.05	-.05	-.08	-.08	-.04	-.04	0
14	+	-	-	-	-	-	-	-	-	0
15	+	-	-	-	-	-	-	-	-	+
16	+	-	-	-	-	-	-	-	-	0
17	+	-	-	-	-	-	-	-	-	-
18	.03	-	-.06	-.38	-.05	-.09	-.10	-.04	-.04	-
19	-	-	-.04	-.05	-.04	-.07	-.08	-.03	-.04	-
20	.03	-	-.06	-.06	-.05	-.09	-.09	-.04	-.04	-
21	.03	-	-.06	-.07	-.05	-.10	-.10	-.04	-.05	-
22	+	-	-.04	-.04	-.04	-.06	-.07	-	-	-
23	.03	-	-.05	-.06	-.04	-.08	-.09	-.03	-.04	-
24	+	-	-.05	-.06	-.05	-.08	-.08	-.03	-.04	-

\+ = positive difference but less than .03;
\- = negative difference but large than -.03;
NA = not available, all power value are equal.